

DOCUMENT RESUME

ED 110 490

TM 004 763

AUTHOR Brandenburg, Dale C.
 TITLE The General Concept of Validity Applied to Student Ratings: Or, Please, General Custer, What Are We Doing Here?
 PUB DATE [Apr 75]
 NOTE 6p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)
 EDRS PRICE MF-\$0.76 HC-\$1.58 PLUS POSTAGE
 DESCRIPTORS Effective Teaching; Higher Education; Instructional Improvement; Student Attitudes; Student Evaluation; *Teacher Evaluation; *Teacher Rating; *Test Validity; Validity

ABSTRACT

The validity of student ratings of instructors and instruction is discussed. The use of student ratings of teachers is increasing, especially in the areas of rank, pay, and retention decisions. Although the author feels that student input is a valid source of information, the type of data that is being collected is based solely on empirical data. The results of this data may also be affected by expected grades, rank of instructor, peer ratings, or required courses as opposed to electives. What is being measured by this type of data must be assessed so that we have a more complete evaluation of teachers and/or students. (DEP)

 * Documents acquired by ERIC include many informal unpublished *
 * materials not available from other sources. ERIC makes every effort *
 * to obtain the best copy available. nevertheless, items of marginal *
 * reproducibility are often encountered and this affects the quality *
 * of the microfiche and hardcopy reproductions ERIC makes available *
 * via the ERIC Document Reproduction Service (EDRS). EDRS is not *
 * responsible for the quality of the original document. Reproductions *
 * supplied by EDRS are the best that can be made from the original. *

THE GENERAL CONCEPT OF VALIDITY
APPLIED TO STUDENT RATINGS

OR

PLEASE, GENERAL CUSTER,¹
WHAT ARE WE DOING HERE?

Dale C. Brandenburg

University of Illinois at Urbana-Champaign

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

Sooner or later when one examines the validity of a given measuring instrument, the question of "valid for what purpose?" arises. Validity of student ratings of instructors and instruction is no exception. As specified in the new issue of *Standards for Educational and Psychological Tests* (American Psychological Association, 1974): "Questions of validity are questions of what may be properly inferred from a test score; validity refers to the appropriateness of inference from test scores or other forms of assessment" (p. 25), and later, "It is important to note that validity is itself inferred, not measured."

To get to my point rather directly, can we infer from our present knowledge and technology based on student ratings that one instructor is a more effective teacher in a given subject than another, as we can infer from our present knowledge and technology based on achievement tests that one examinee possesses more knowledge of a given topic than another examinee? I think not; and yet the use of student ratings for rank, pay, and retention decisions is becoming more frequent. This administrative use of student ratings is undoubtedly the weakest area for the validity of student ratings from a technical viewpoint.

Before I return to this point, I want to talk about two other general purposes that involve the use of student ratings: Improvement of instruction and information to guide the student in selection of courses. The latter purpose of student ratings is probably the area less susceptible to validity concerns than the former. Published student ratings undoubtedly provide students with the most solid and reliable piece of information they have for the selection of instructors and courses, given obviously the choice of such.

However, when one considers the other two purposes of student ratings, improvement of instruction and personnel decisions, we must admit that validity data are weak, contradictory, and consequently they do not lead us

¹Part of a symposium, *Conceptualizations of validity for student ratings of instruction*, presented at the AERA Annual Meeting, April 2, 1975.

to discover any convincing definition for an effective teacher. In the meantime, we advocate to new users developing faculty evaluation systems that student input is a necessary, valid source of information.

I agree that student data are valid input, but let's look at the type of data we are collecting. Most items on student rating questionnaires are concerned with process-type variables. Recent approaches by Hoyt (1969), Hogan and Hartley (1972), and Jaeger and Feijo (1974) have begun to focus on outcomes (more appropriately, student-viewed outcomes). Which type of variable is more important for a given purpose: Process or outcome? We can argue more about this later; but for the moment, let's consider what we are doing. Most questionnaires, to my knowledge, have been formulated from empirical bases; other instruments may have begun by asking certain groups of individuals for their definition of an effective teacher. I suggest that this latter approach is for the most part also empirical, i.e., based on past experience and perception.

How many student rating instruments do you know that have been based on something called a theory of instruction? I don't know of any. It seems to me that we have relied upon something based entirely on empiricism, and we have proceeded for a long time to draw inferences from our empirical data base. The consequences have not, thus far, been chaotic; but I don't believe we know where we're heading. It reminds me of what Suppés (1974, p. 6) said here last year: "Reliance on bare empiricism or bare intuition in educational practice is a mental form of streaking, and nudity of mind is not as appealing as nudity of body."

To illustrate my point, a recent careful, comprehensive review of the teacher evaluation literature since 1929, (Batista, *Characteristics of effective college teachers*, unpublished manuscript, December, 1974) found that the following four general types of items appear in almost every student rating form:

1. Subject matter: Knowledge of subject matter and enthusiasm for it.
2. Classroom performance: Organization of lectures, ability to communicate a clearness of presentation.
3. Relations with students: Awareness of student needs and interests, acceptance of students' points of view and facilitation of student participation.
4. Student learning: Stimulation of students' thinking, capability of promoting students' interest in the field and facilitation of students' personal and professional growth.

It is not difficult to discern that what we have above can generally be thought of as process variable. It was Hogan and Hartley's paper (1972) that got us thinking more about outcomes. What had we been doing previously--the GIGO principle must have been working well in our factor analytic studies of student ratings (garbage in, garbage out). Why is it that this factor was undiscovered, or could we have been relying too heavily on empiricism?

Still I can agree that our present best single piece of information in the evaluation of instructor performance, i.e., the teaching component, is student ratings of performance. We must realize, however, that student ratings are an indirect evaluation of teaching. Direct evaluation (or primary evaluation as Scriven [unpublished manuscript, 1974] defines it) is "determination of the gains in understanding or learning or attitudes resulting from a particular exposure to teaching, possibly combined with some direct checks on the justice and pleasure of the classroom process" (p. 9). About indirect or secondary evaluation, Scriven states: "One needs to face the fact that almost all secondary evaluation (of teaching) is extremely unreliable [defined as the ability to identify a superior teacher] and that a great deal of it is not only completely worthless, but so unreliable that its use does more harm than good" (p. 9). Scriven also complains that at present there exists no "defensible global (synthesized) rating of teaching merit" (p. 10). To complete the argument on secondary evaluation, Scriven states: "The basic axiom here is, of course, that it is logically impossible for any indications (and hence any secondary evaluation) to be validated unless some primary evaluation is done" (p. 10). We know that there exists some studies which have shown that student ratings are related to student learning or achievement, but here again the results are contradictory. If we forget about Rodin and Rodin's study, and I am inclined to do that, we are still left hanging a bit with the present studies. The reason for this is that none of these studies have utilized the normal college classroom as the unit of analysis--I argue for this because I do not consider teaching assistants and their classes to be reflective of the mainstream of college instruction.

In the total scheme for the evaluation of teaching, there is a place for student ratings. Even Scriven (1974, p. 11) agrees that student ratings should be given some weight because "there is a commonsensical and moral basis for expecting them to be at least weakly correlated to success as a teacher." In order to move forward on research in this area (to borrow Scriven's terms) we must distinguish between indicator variables (factors that are correlated with good teaching) and improvement variables (factors that can be emulated and that control teaching merit). In the review of literature cited previously (Batista, 1974) he formulates a definition of an effective teacher from a composite of student rating results. It is an instructor who is "cultural and/or knowledgeable, enthusiastic about his subject area, emotionally stable, concerned for the welfare of his students, able to communicate his ideas clearly, and capable of stimulating students' learning and growth" (p. 64). If we believe that our clients, the college faculty at large, should try to emulate the current profiles of student rating results of our best teachers we may be "at best fruitless and possibly counter productive" (Scriven, p. 22) in trying to improve instruction. This, then, is the primary reason why we should not include process items on questionnaires whose results are going to be fed into personnel decisions.

Let us consider the following questions:

1. Are student ratings related to colleague ratings?
2. Do students in required courses rate differently than students in elective courses?

3. When is the proper time to administer ratings?
4. Do present student ratings correlate with alumni ratings?
5. Do student ratings on one form correlate with those from another?
6. Do ratings differ depending on the rank of the instructor?
7. Do ratings differ among subject matter areas?
8. Are student ratings related to cognitive achievement?
9. How much does expected grade influence the ratings?

While I do not deny the importance of some of these questions, they are concerned with one exception (#8) with indirect evaluation and none have to do with improvement variables. Yet, we advocate the use of student ratings for improvement of instruction and administrative decisions. Are we improving instruction or are we really improving the ratings? Are we promoting better instructors or instructors with better ratings? What standard do we have to go by, i.e., can we tell an instructor that if he does this he will be a better instructor and to hell with what the ratings say? In general, I'll have to agree with Menne (*NCME Measurement News*, April, 1974), who stated that what we are measuring is teacher performance and not teacher effectiveness. It appears that Dr. Fox is a grand example of where we should not be going.

To return to the introduction, are our inferences from these measures appropriate, i.e., valid, or are we perpetrating circular empiricisms? I believe that there is some room for debate.

References

- American Psychological Association. *Standards for educational and psychological tests*. Washington, D.C., 1974.
- Batista, E. E. *Characteristics of effective college teachers*. Unpublished manuscript, Measurement and Research Division, Office of Instructional Resources, University of Illinois at Urbana-Champaign, 1974.
- Hogan, T. & Hartley, E. Some additional factors in student evaluations of courses. *American Educational Research Journal*, 1972, 9, 241-250.
- Hoyt, D. P. *Student reactions to instruction and courses*. Unpublished manuscript, Office of Educational Resources, Kansas State University, 1969.
- Jaeger, R. M. & Feijo, T. D. Some psychometric questions in the evaluation of professors. *Journal of Educational Psychology*, 1974, 66, 416-423.
- Menne, J. W. Teacher evaluation: Performance or effectiveness? *NCME Measurement News*, April, 1974, 11-12.
- Scriven, M. *The evaluation of teaching*. Unpublished manuscript, University of California at Berkeley, 1974.
- Suppes, P. *The place of theory in educational research*. Presidential address to the American Educational Research Association, April 17, 1974. *Educational Researcher*, 1974, 3, 3-10.